JOURNAL OF MULTIVARIATE ANALYSIS 42, 284–298 (1992)

Constrained Kantorovich Inequalities and Relative Efficiency of Least Squares*

SONG-GUI WANG

Department of Mathematics, University of Science and Technology of China, Hefei, 230026 People's Republic of China

AND

JUN SHAO

Department of Mathematics, University of Ottawa, Ottawa, K1N 6N5 Canada,

Communicated by the Editors

This paper establishes a type of Kantorovich inequality subject to some constraints and obtains some lower bounds for the relative efficiency of the least squares. These lower bounds can be much sharper than that obtained by using the unconstrained Kantorovich inequality. Multivariate extensions of the results are also obtained. Some interesting examples are presented.

1. INTRODUCTION

Let A be an $n \times n$ symmetric positive definite matrix and x be an $n \times 1$ vector satisfying $x'x = 1$ (all matrices and vectors considered in this article have real elements). The Kantorovich inequality asserts that

$$x'Axx'A^{-1}x \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}, \tag{1.1}$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ are the eigenvalues of A. In the last two decades, much attention has been focused on multivariate extension of (1.1). Let X be an

Received September 4, 1991; revised January 16, 1992.

AMS 1980 subject classifications: primary 62J05; secondary 62F11, 15A42.

Key words and phrases: Linear models, linear constraints, relative efficiency, BLUE.

* Both authors were supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Song-Gui Wang was also supported by grants from Faculty Development Fund at University of Ottawa and from the National Natural Sciences Foundation of China.

0047-259X/92 $5.00

$n \times p$ matrix such that $X'X = I_p$, the identity matrix of order p. Bloomfield and Watson [3] and Knott [6] showed that

$$|X'AX|\,|X'A^{-1}X| \leqslant \prod_{i=1}^{\min(n,n-p)} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i \lambda_{n-i+1}}. \tag{1.2}$$

Khatri and Rao [4] further showed that for $n \times p$ matrices X and Y satisfying $X'X = Y'Y = I_p$, $|X'AY|\,|Y'A^{-1}X|$ has also the upper bound given by (1.2). Other extensions can be found in Khatri and Rao [5] and Rao [10].

A recent note by Marshall and Olkin [9] provided a different multivariate extension of (1.1): They showed that for $X'X = I_p$,

$$X'A^{-1}X \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n} (X'AX)^{-1}, \tag{1.3}$$

where $A \leqslant B$ means $B - A$ is nonnegative definite for two symmetric matrices A and B. Note that (1.3) is the same as (1.1) if $p = 1$. Baksalary and Puntanen [1] extended (1.3) by allowing A to be singular and, consequently, replacing A^{-1} and λ_n by the Moore–Penrose inverse A^+ and the smallest nonzero eigenvalue of A, respectively.

The Kantorovich inequality and its extensions play important roles in the study of relative efficiency of the least squares estimator (LSE) in linear models (e.g., Bloomfield and Watson [3]; Khatri and Rao [4]). Note that the LSE is a linear unbiased estimator but not necessarily the best linear unbiased estimator (BLEU). However, because of its simplicity and popularity, the LSE may still be preferred unless its relative efficiency (to the BLEU) is substantially low. Thus, some lower bounds for the relative efficiency of the LSE are useful in studying the performance of the LSE. The Kantorovich inequality provides an important tool in establishing such bounds.

However, in efficiency studies the lower bounds produced by the inequalities (1.1) and (1.3) are often not sharp enough, since only the smallest and the largest eigenvalues are involved in (1.1) and (1.3). Note that inequality (1.2) involves not just λ_1 and λ_n and therefore provides more information in comparing the efficiencies of the LSE and BLEU (see Khatri and Rao [4]). It is desired to obtain improved inequalities of type (1.1) or (1.3) in terms of other eigenvalues of A, but improving (1.1) or (1.3) cannot be done without putting any constraints on x (or on X), since the equality in (1.1) or (1.3) can be achieved by choosing a particular x or X. The main purpose of this article is to establish some Kantorovich inequalities subject to linear constraints and to study their applications in studying the efficiency of the LSE.

In Section 2, we establish a constrained Kantorovich inequality and obtain some lower bounds for the relative efficiency of the LSE. These lower bounds can be much sharper than that obtained by using the unconstrained Kantorovich inequality (Magness and McGuire [8]). Several examples of applications of our main results are discussed.

Section 3 presents some matrix versions of the constrained Kantorovich inequalities, which are multivariate extensions of the result in Section 2. These algebraic results are established by using the estimation theory in a Gauss–Markov model, an approach different from that of Marshall and Olkin [9]. Some examples of applications of the results are also presented.

Although we focus on constrained inequalities of type (1.1) or (1.3), constrained Kantorovich inequalities of other types, such as (1.2), can be established in a similar manner.

2. The Constrained Kantorovich Inequality and Its Applications

The following result is an extension of (1.1). It will be called the constrained Kantorovich inequality.

Lemma 1. *Let A be an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and the corresponding orthonormal eigenvectors $\varphi_1, ..., \varphi_n$. Let i_l, $l = 1, ..., k$, be integers satisfying $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$, $k \leqslant n$, $\Phi_1 = (\varphi_{i_1}, ..., \varphi_{i_k})$, and $\mathscr{M}(\Phi_1)$ be the linear space generated by the columns of Φ_1. Then*

$$\sup_{\substack{x \in \mathscr{M}(\Phi_1) \\ x \neq 0}} \frac{x'Axx'A^{-1}x}{(x'x)^2} = \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} \tag{2.1}$$

and the maximum is attained if and only if $x = c(\varphi_{i_1} \pm \varphi_{i_k})$ with a constant c.

Proof. Without loss of generality, we assume that $i_l = l$, $l = 1, ..., k$. Denote $\Phi_1 = (\varphi_1, ..., \varphi_k)$, and $\Lambda_1 = \mathrm{diag}(\lambda_1, ..., \lambda_k)$. Since $x \in \mathscr{M}(\Phi_1)$ if and only if there is a $k \times 1$ vector t *such that* $x = \Phi_1 t$, we have

$$\sup_{\substack{x \in \mathscr{M}(\Phi_1) \\ x \neq 0}} \frac{x'Axx'A^{-1}x}{(x'x)^2} = \sup_{t \neq 0} \frac{t'\Lambda_1 tt'\Lambda_1^{-1}t}{(t't)^2} = \frac{(\lambda_1 + \lambda_k)^2}{4\lambda_1\lambda_k}$$

by using (1.1). Thus (2.1) is proved.

The remaining part follows from the proof of (1.2) in Bloomfield and Watson [3, pp. 123–124]. ∎

Note that the supremum in (2.1) is taken subject to the constraint $x \in \mathcal{M}(\Phi_1)$. If $\Phi_1 = (\varphi_1, ..., \varphi_n)$, then $\mathcal{M}(\Phi_1)$ is the whole space and (2.1) reduces to the unconstrained Kantorovich inequality (1.1).

We now apply (2.1) to obtain lower bounds for the relative efficiency of the LSE in linear models. Consider the general linear model

$$y = X\beta + \varepsilon, \qquad E(\varepsilon) = 0. \qquad \mathrm{Cov}(\varepsilon) = V, \tag{2.2}$$

where y is an $n \times 1$ vector of observations, X is an $n \times p$ design matrix, β is a $p \times 1$ vector of unknown parameters, ε is an $n \times 1$ vector of random errors, and V is positive definite. For any estimable function $c'\beta$, its LSE is $c'\hat{\beta} = c'(X'X)^- X'y$, where A^- stands for a generalized inverse of A, and the "BLEU" of $c'\beta$ is $c'\tilde{\beta} = c'(X'V^{-1}X)^- X'V^{-1}y$. The relative efficiency of $c'\hat{\beta}$ with respect to $c'\tilde{\beta}$ is

$$\mathrm{Eff}(c'\hat{\beta}) = \frac{\mathrm{Var}(c'\tilde{\beta})}{\mathrm{Var}(c'\hat{\beta})}. \tag{2.3}$$

It is well known that $\mathrm{Eff}(c'\hat{\beta}) \leqslant 1$ and under some conditions, $c'\hat{\beta} = c'\tilde{\beta}$ for any estimable function $c'\beta$ and therefore $\mathrm{Eff}(c'\hat{\beta}) = 1$ (see, e.g., Rao and Mitra [11, p. 155]; Zyskind [12]). In general, $\mathrm{Eff}(c'\hat{\beta})$ can be smaller than 1 and a lower bound for $\mathrm{Eff}(c'\hat{\beta})$, due to Magness and McGuire [8], is

$$\mathrm{Eff}(c'\hat{\beta}) \geqslant \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}, \tag{2.4}$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ are the eigenvalues of V. The lower bound in (2.4) is obtained by using the Kantorovich inequality (1.1) and it holds for all X and all estimable functions $c'\beta$. However, it is quite common to have some information about X and therefore the lower bound in (2.4) can be improved. The following result gives a sharper lower bound of $\mathrm{Eff}(c'\hat{\beta})$ by using Lemma 1.

THEOREM 1. *Assume model* (2.2). *Let* $r = \mathrm{rank}(X)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ *be the eigenvalues of* V *and* $\varphi_1, ..., \varphi_n$ *be the corresponding orthonormal eigenvectors. Suppose that* $\mathcal{M}(X) \subset \mathcal{M}(\varphi_{i_1}, ..., \varphi_{i_k})$ *for some integers* i_l *satisfying* $1 \leqslant i_l \leqslant \cdots \leqslant i_k \leqslant n$, $k \leqslant n$. *Then for any estimable* $c'\beta$,

$$\mathrm{Eff}(c'\hat{\beta}) = 1 \qquad \text{if } r = k, \tag{2.5}$$

$$\geqslant \frac{4\lambda_{i_1}\lambda_{i_k}}{(\lambda_{i_1} + \lambda_{i_k})^2} \qquad \text{if } r < k. \tag{2.6}$$

Proof. If $r = k$, then $\mathcal{M}(X) = \mathcal{M}(\varphi_{i_1}, ..., \varphi_{i_k})$. From Theorem 2 of

Zyskind [12], $c'\hat{\beta}=c'\tilde{\beta}$ for any estimable function $c'\beta$ and therefore (2.5) holds.

Suppose that $r<k$. Consider the full rank decomposition $X=PQ'$, where the columns of P form an orthonormal basis for $\mathcal{M}(X)$. From the estimability of $c'\beta$, there exists a $p\times 1$ vector α such that $c=X'\alpha$. Let $w=P'\alpha$. Then

$$\operatorname{Var}(c'\hat{\beta})=\alpha'PP'VPP'\alpha=w'PVP'w \tag{2.7}$$

and

$$\operatorname{Var}(c'\tilde{\beta})=\alpha'X(X'V^{-1}X)^{-}\,X'\alpha=w'(P'V^{-1}P)^{-1}\,w\geqslant\frac{(w'w)^2}{w'P'V^{-1}Pw} \tag{2.8}$$

by the Cauchy–Schwarz inequality. Combining (2.7)–(2.8), we obtain that

$$\operatorname{Eff}(c'\hat{\beta})=\frac{(w'w)^2}{w'P'VPww'P'V^{-1}Pw}=\frac{(u'u)^2}{u'Vuu'V^{-1}u}\geqslant\frac{4\lambda_{i_1}\lambda_{i_k}}{(\lambda_{i_1}+\lambda_{i_k})^2},$$

where $u=Pw$ and the last inequality follows from Lemma 1 since $u=Pw=PP'\alpha\in\mathcal{M}(X)\subset\mathcal{M}(\varphi_{i_1},...,\varphi_{i_k})$. This completes the proof. ∎

Remarks. (i) Since $\mathcal{M}(X)\subset\mathcal{M}(\varphi_1,...,\varphi_n)$, result (2.4) is a special case of (2.6).

(ii) The best lower bound in (2.6) corresponds to the smallest eigen-subspace containing $\mathcal{M}(X)$. If $\mathcal{M}(\varphi_{i_1},...,\varphi_{i_k})$ is the smallest eigen-subspace containing $\mathcal{M}(X)$, then the lower bound of $\operatorname{Eff}(c'\hat{\beta})$ is unrelated to λ_i with $i\notin\{i_1,...,i_k\}$.

(iii) If $r<k$, then either $\lambda_{i_1}<\lambda_{i_k}$ or $\mathcal{M}(X)$ is generated by eigenvectors of V. This means if $r<k$, then the lower bound in (2.6) cannot be 1 unless the columns of X are eigenvectors of V. This assertion can be shown as follows. If $\lambda_{i_1}=\lambda_{i_k}$, then $\lambda_{i_1}=\cdots=\lambda_{i_k}$ and $\mathcal{M}(\varphi_{i_1},...,\varphi_{i_k})$ is an eigen-subspace corresponding to λ_{i_1}. From $\mathcal{M}(X)\subset\mathcal{M}(\varphi_{i_1},...,\varphi_{i_k})$, any basis of the subspace $\mathcal{M}(X)$ consists of eigenvectors of V corresponding to λ_{i_1}.

The lower bound (2.6) can be much sharper than the lower bound (2.4), as the following examples indicate.

EXAMPLE 1. Consider the following special model of (2.2):

$$y=\mathbf{1}_n\theta+\varepsilon,\qquad E(\varepsilon)=0,\qquad \operatorname{Cov}(\varepsilon)=V,$$

where $\mathbf{1}_n$ is the $n\times 1$ vector of ones and θ is an unknown parameter. Blom [2] and Zyskind [13, p. 1361] observed that the LSE $\bar{y}=\sum_{i=1}^n y_i/n$ is also the BLEU if and only if V has all of its row totals equal, because $V\mathbf{1}_n=c\mathbf{1}_n$

for some scalar c implies $X=\mathbf{1}_n$ is an eigenvector of V. An example of such a covariance matrix V for $n=4$ is the following:

$$V_0=\begin{pmatrix} \lambda/2+2 & \lambda/2-2 & 0 & 0 \\ \lambda/2-2 & \lambda/2+2 & 0 & 0 \\ 0 & 0 & \lambda/2+1 & \lambda/2-1 \\ 0 & 0 & \lambda/2-1 & \lambda/2+1 \end{pmatrix}.$$

For this case, $V_0\mathbf{1}_4=\lambda\mathbf{1}_4$ and $\bar{y}$ is the BLUE. However, if V_0 is perturbed by a small change,

$$V_\varepsilon=\begin{pmatrix} \lambda/2+2 & \lambda/2-2 & 0 & 0 \\ \lambda/2-2 & \lambda/2+2 & 0 & 0 \\ 0 & 0 & (\lambda+\varepsilon)/2+1 & (\lambda+\varepsilon)/2-1 \\ 0 & 0 & (\lambda+\varepsilon)/2-1 & (\lambda+\varepsilon)/2+1 \end{pmatrix},$$

then $\mathbf{1}_4$ is not an eigenvector of V_ε when $\varepsilon>0$ and therefore $\bar{y}$ is not the BLEU. The matrix V_ε has four eigenvalues: λ, $\lambda+\varepsilon$, 2, and 4. The eigenvectors corresponding to λ and $\lambda+\varepsilon$, respectively, are $\varphi_1=((1/\sqrt{2})\,\mathbf{1}_2', 0)'$ and $\varphi_2=((0, 1/\sqrt{2})\,\mathbf{1}_2')'$. Since $\mathcal{M}(\varphi_1, \varphi_2)$ is the smallest eigen-subspace containing $\mathbf{1}_4$, it follows from Theorem 1 that

$$\operatorname{Eff}(\bar{y})\geqslant\frac{4\lambda(\lambda+\varepsilon)}{(2\lambda+\varepsilon)^2}.$$

Note that $\operatorname{Eff}(\bar{y})\to 1$ as $\varepsilon\to 0$. Hence the LSE $\bar{y}$ is quite robust against small perturbations in this problem.

If we use the lower bound in (2.4), then by assuming $\lambda+\varepsilon<2$ as an example, we obtain

$$\operatorname{Eff}(\bar{y})\geqslant\frac{16\lambda}{(\lambda+4)^2}.$$

Thus, in this example, the lower bound in (2.6) is much sharper than that in (2.4). In fact one cannot show the robustness of $\bar{y}$ by using the lower bound in (2.4).

EXAMPLE 2. Consider the random effect model

$$y=\mathbf{1}_n\mu+U\xi+\varepsilon, \qquad E(\varepsilon)=0, \qquad \operatorname{Cov}(\varepsilon)=\sigma_\varepsilon^2 I_n,$$

where y is an $n\times 1$ vector of observations, μ is a non-random unknown parameter, U is an $n \times p$ known matrix of full rank, ξ is a $p \times 1$ random vector with $E(\xi)=0$ and $\text{Cov}(\xi)=\sigma_\xi^2 I_n$, and ε is an $n \times 1$ vector of errors and is independent of ξ. Let

$$U = \Phi \begin{pmatrix} \Lambda^{1/2} \\ 0 \end{pmatrix} Q'$$

be the singular value decomposition, where $\Phi = (\varphi_1, ..., \varphi_n)$ with $\Phi'\Phi = I_n$, $\Lambda = \text{diag}(\lambda_1, ..., \lambda_p)$ with $\lambda_1 \geqslant \cdots \geqslant \lambda_p > 0$ and Q is a $p \times p$ matrix with $Q'Q = I_p$. It is easy to verify that $V = \text{Cov}(y) = \sigma_\xi^2 UU' + \sigma_\varepsilon^2 I_n$ has eigenvalues $\sigma_\xi^2 \lambda_1 + \sigma_\varepsilon^2, ..., \sigma_\xi^2 \lambda_p + \sigma_\varepsilon^2, \sigma_\varepsilon^2$ (multiplicity is $n-p$) with the corresponding orthonormal eigenvectors $\varphi_1, ..., \varphi_n$. We consider the following two cases:

(1) If all the row totals of UU' are equal, then the LSE $\bar{y}$ is the BLEU of μ.

(2) If the row totals of UU' are not equal, then $\mathbf{1}_n$ is equal to neither φ_i, nor linear combinations of $\varphi_{p+1}, ..., \varphi_n$. Assume that $\mathbf{1}_n = \sum_{i \in I} \alpha_i \varphi_i$ for a subset $I \subset \{1, ..., n\}$ and some constants α_i. Let $i_0 = \min\{i : i \in I\}$. Then by Theorem 1,

$$\text{Eff}(\bar{y}) \geqslant \frac{4(1 + \lambda_{i_0} \sigma_\xi^2/\sigma_\varepsilon^2)}{(2 + \lambda_{i_0} \sigma_\xi^2/\sigma_\varepsilon^2)^2}.$$

This lower bound is better than the one obtained from (2.4).

Let $e = (e_1, ..., e_n)'$ be a random vector satisfying

$$E(e_i) = 0, \qquad \text{Cov}(e_i, e_j) = \begin{cases} \sigma^2 & i = j, \\ \sigma^2 \rho & i \neq j. \end{cases} \tag{2.9}$$

Then the covariance matrix $\text{Cov}(e)$ has the following property: It has a single eigenvalue $\lambda = \sigma^2[1 + (n-1)\rho]$ with eigenvector $\mathbf{1}_n$ and a repeated eigenvalue $\tau = \sigma^2(1-\rho)$ with multiplicity $n-1$. More generally, it is very common that the covariance matrix V in the linear model (2.2) has the form

$$V = \text{block.diag}(V_{n_1}, ..., V_{n_k}), \tag{2.10}$$

where each V_{n_i} is a positive definite matrix of order n_i and satisfies

$$V_{n_i} \text{ has a single eigenvalue } \lambda_i \text{ with eigenvector } \mathbf{1}_{n_i}; \tag{2.11}$$

$$V_{n_i} \text{ has a repeated eigenvalue } \tau_i \text{ with multiplicity } n_i - 1. \tag{2.12}$$

We have the following result as a consequence of Theorem 1.

COROLLARY 1. *Suppose that the covariance matrix V in model* (2.2) *satisfies* (2.10)–(2.12). *Then we have the following results*:

(i) *Let* $J_1=(\mathbf{1}'_{n_1}, 0, ..., 0)'$, $J_2=(0, \mathbf{1}'_{n_2}, ..., 0)'$, ..., $J_k=(0, 0, ..., \mathbf{1}'_{n_k})'$. *If* $\mathcal{M}(X)\subset\mathcal{M}(J_1, ..., J_k)$, *then for any estimable function* $c'\beta$,

$$\mathrm{Eff}(c'\hat{\beta})\geqslant\frac{4\lambda_*\lambda^*}{(\lambda_*+\lambda^*)^2},$$

where $\lambda_*=\min\{\lambda_i\}$ *and* $\lambda^*=\max\{\lambda_i\}$. *If in addition*, $\lambda_*=\lambda^*$, *then* $c'\hat{\beta}$ *is the* BLUE *for any estimable* $c'\beta$.

(ii) *If* $X'J_i=0$ *for all* i, *then for any estimable function* $c'\beta$,

$$\mathrm{Eff}(c'\hat{\beta})\geqslant\frac{4\tau_*\tau^*}{(\tau_*+\tau^*)^2},$$

where $\tau_*=\min\{\tau_i\}$ *and* $\tau^*=\max\{\tau_i\}$. *If in addition*, $\tau_*=\tau^*$, *then* $c'\hat{\beta}$ *is the* BLUE *for any estimable* $c'\beta$.

(iii) *If in* (2.10), $k=1$ *and* $\mathbf{1}_n$ *is a column of* X, *then* $c'\hat{\beta}$ *is the* BLUE *for any estimable* $c'\beta$.

Proof. (i) For each i, J_i is an eigenvector of V corresponding to the eigenvalue λ_i. Hence the result follows from Theorem 1.

(ii) $X'J_i=0$ for all i implies that $\mathcal{M}(X)$ is contained in the subspace generated by the eigenvectors of V corresponding to the eigenvalues $\tau_1, ..., \tau_n$. Hence the result follows from Theorem 1.

(iii) Since $\mathrm{rank}(X)=r$ and $\mathbf{1}_n$ is a column of X, we can find linearly independent vectors $z_1, ..., z_r$ such that $z_1=\mathbf{1}_n$ and $\mathcal{M}(X)=\mathcal{M}(z_1, ..., z_r)$. Condition (2.11) and $k=1$ implies that $\mathbf{1}_n$ is the eigenvector corresponding to the single eigenvalue of V. Let $\varphi_1=\mathbf{1}_n, \varphi_2, ..., \varphi_n$ be orthogonal eigenvectors of V. Then $\mathcal{M}(X)\subset\mathcal{M}(\varphi_1, ..., \varphi_n)$ and there are constants c_{ij} such that

$$z_j=\sum_{i=1}^{n} c_{ij}\varphi_i=c_{1j}\mathbf{1}_n+\sum_{i=1}^{n} c_{ij}\varphi_i, \qquad j=1, ..., r.$$

Let $\xi_j=\sum_{i=2}^{n} c_{ij}\varphi_i, j=2, ..., r$. Then $\mathcal{M}(X)\subset\mathcal{M}(\mathbf{1}_n, \xi_2, ..., \xi_r)$. From (2.12), ξ_j are eigenvectors of V. Hence the result follows from Theorem 1. ∎

EXAMPLE 3. A simple example where $\mathcal{M}(X)\subset\mathcal{M}(J_1, ..., J_k)$ is the one-way analysis of variance model:

$$y_{ij}=\mu+\alpha_i+\varepsilon_{ij}, \qquad j=1, ..., n_i, \qquad i=1, ..., k.$$

If for $i \neq i'$, $\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$ and $\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{ij'})$ is given by (2.9) with $\sigma^2 = \sigma_i^2$, $\rho = \rho_i$, then conditions (2.10)–(2.12) are satisfied. Hence Corollary 1 can be applied to this problem.

In some problems we are only interested in parameter functions of the form $c'\beta$ for some c. The following corollary follows directly from Theorem 1:

COROLLARY 2. *Suppose that* $\mathcal{M}(X(X'X)^+ C) \subset \mathcal{M}(\varphi_{i_1}, ..., \varphi_{i_k})$ *for some* $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$, *where* $(X'X)^+$ *is the Moore–Penrose inverse of* $X'X$, $C = (c_1, ..., c_l)$ *and* $c_i'\beta$, $i = 1, ..., l$, *are linearly independent estimable functions. Then for any* $c \in \mathcal{M}(C)$,

$$\mathrm{Eff}(c'\hat{\beta}) \geqslant \frac{4\lambda_{i_1}\lambda_{i_k}}{(\lambda_{i_1} + \lambda_{i_k})^2}.$$

EXAMPLE 4. Consider the analysis of variance model,

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \qquad i = 1, ..., a, \qquad j = 1, ..., b, \tag{2.13}$$

where α_i are effects of a factor, $\sum_{i=1}^{a} \alpha_i = 0$, and β_j are block effects, $\sum_{j=1}^{b} \beta_j = 0$. Between blocks, the errors ε_{ij} are independent. Within the jth block, $\mathrm{Cov}(\varepsilon_{jl}, \varepsilon_{jk}) = \sigma_j^2 \rho_j$ if $l \neq k$ and $= \sigma_j^2$ if $l = k$. Thus, the error covariance matrix V satisfies (2.10)–(2.12).

We now consider the LSE of any contrast of $\alpha_1, ..., \alpha_a$. Let $c_i = (0, 1, 0, ..., 0, -1, 0, ..., 0)'$ be an $ab \times 1$ vector, where -1 is in the $(i+2)$th component of c_i, $i = 1, ..., a-1$, and $C = (c_1, ..., c_{a-1})$. Let $\gamma = (\mu, \alpha_1, ..., \alpha_a, \beta_1, ..., \beta_b)'$ be the parameter vector. Since $c_i'\gamma$, $i = 1, ..., a-1$, are linearly independent contrasts, any contrasts of $\alpha_1, ..., \alpha_a$ can be expressed as a linear combination of $c_i'\gamma$, $i = 1, ..., a-1$. Hence we are interested in estimating functions of the form $c'\gamma$, $c \in \mathcal{M}(C)$.

To apply Corollary 2, we need to calculate the matrix $X(X'X)^+ C$, where X is the design matrix of model (2.13). Let $z_i = (0, 1, 0, ..., 0, -1, 0, ..., 0)'$ be an $ab \times 1$ vector, where -1 is in the $(ib+1)$th component, $i = 1, ..., a-1$. Then $C = X'Z$, $Z = (z_1, ..., z_{a-1})$. Using the standard formula for evaluating the LSE of α_i and β_j, we obtain that

$$X(X'X)^+ C = (h_1, ..., h_{a-1}),$$

where

$$\begin{aligned} h_1 &= (\mathbf{1}_b', -\mathbf{1}_b', 0, ..., 0)', \\ h_2 &= (\mathbf{1}_b', 0, -\mathbf{1}_b', ..., 0)', \\ &\vdots \quad \vdots \\ h_{a-1} &= (\mathbf{1}_b', 0, 0, ..., -\mathbf{1}_b')'. \end{aligned}$$

Note that

$$\begin{aligned} J_1 &= (\mathbf{1}'_b, 0, ..., 0)', \\ J_2 &= (0, \mathbf{1}'_b, ..., 0)', \\ &\vdots \\ J_a &= (0, 0, ..., \mathbf{1}'_b)', \end{aligned}$$

are orthogonal eigenvectors of V corresponding to the single eigenvalues $\lambda_1 = \sigma_1^2[1+(b-1)\rho_1]$, ..., $\lambda_a = \sigma_a^2[1+(b-1)\rho_a]$, respectively. Apparently,

$$\mathcal{M}(X(X'X)^+ C) \subset \mathcal{M}(J_1, ..., J_a)$$

and therefore by Corollary 2,

$$\mathrm{Eff}(c'\hat{\gamma}) \geqslant \frac{4\lambda_* \lambda^*}{(\lambda_* + \lambda^*)^2}, \tag{2.14}$$

for any $c \in \mathcal{M}(C)$, where $\lambda_* = \min\{\lambda_1, ..., \lambda_a\}$ and $\lambda^* = \max\{\lambda_1, ..., \lambda_a\}$. In particular, if $\sigma_i^2 \equiv \sigma^2$ and $\rho_i \equiv \rho$, then $\mathrm{Eff}(c'\hat{\gamma}) = 1$, i.e., for any contrast of $\alpha_1, ..., \alpha_a$, its LSE is also the BLUE. The lower bound in (2.14) is much sharper than the lower bound obtained in (2.4), which is (assuming $\rho_i > 0$)

$$\mathrm{Eff}(c'\hat{\gamma}) \geqslant \frac{4\tau_* \lambda^*}{(\tau_* + \lambda^*)^2},$$

where $\tau_* = \min\{\sigma_1^2(1-\rho_1), ..., \sigma_a^2(1-\rho_a)\}$.

3. Matrix Version of the Constrained Kantorovich Inequality

In this section we establish some matrix versions of the constrained Kantorovich inequality. Consider the general linear model

$$y = X\beta + \varepsilon, \qquad E(\varepsilon) = 0, \qquad \mathrm{Cov}(\varepsilon) = A^{-1}, \tag{3.1}$$

where A is an $n \times n$ positive definite matrix. Let B be a positive definite matrix. Premultiplying (3.1) by $B^{1/2}$ gives

$$B^{1/2}y = B^{1/2}X\beta + e, \qquad E(e) = 0, \qquad \mathrm{Cov}(e) = B^{1/2}A^{-1}B^{1/2}, \tag{3.2}$$

where $e = B^{1/2}\varepsilon$. Let $\mu = B^{1/2}X\beta$. The LSE of μ based on model (3.2) is

$$\hat{\mu} = B^{1/2}X(X'BX)^- X'By$$

with covariance matrix

$$\mathrm{Cov}(\hat{\mu}) = B^{1/2}X(X'BX)^{-}\ X'BA^{-1}BX(X'BX)^{-}\ X'B^{1/2}. \tag{3.3}$$

If A and B are known matrices, then the best linear unbiased estimator of μ based on the same model is

$$\tilde{\mu} = B^{1/2}X(X'AX)^{-}\ X'Ay$$

with covariance matrix

$$\mathrm{Cov}(\tilde{\mu}) = B^{1/2}X(X'BX)^{-}\ X'B^{1/2}. \tag{3.4}$$

We now use (3.3)–(3.4) and Theorem 1 to establish the following result:

THEOREM 2. *Let X be an $n \times p$ matrix with rank r, A and B be two $n \times n$ positive definite matrices, $\tau_1 \geqslant \cdots \geqslant \tau_n$ be the eigenvalues of $B^{1/2}A^{-1}B^{1/2}$, and $\xi_1, ..., \xi_n$ be the corresponding orthonormal eigenvectors. Suppose that $\mathscr{M}(B^{1/2}X) \subset \mathscr{M}(\xi_{i_1}, ..., \xi_{i_k})$ for some $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$, $k \leqslant n$. Then*

$$X'BA^{-1}BX \leqslant \frac{(\tau_{i_1} + \tau_{i_k})^2}{4\tau_{i_1}\tau_{i_k}}\ X'BX(X'AX)^{-}\ X'BX. \tag{3.5}$$

If $k = r$, then

$$X'BA^{-1}BX = X'BX(X'AX)^{-}\ X'BX. \tag{3.6}$$

Proof. Consider model (3.2). It follows from Theorem 1 that for any $n \times 1$ vector $c \neq 0$,

$$\begin{aligned} \mathrm{Eff}(c'\hat{\mu}) &= 1 && \text{if } r = k, \\ &\geqslant \frac{4\tau_{i_1}\tau_{i_k}}{(\tau_{i_1} + \tau_{i_k})^2} && \text{if } r < k. \end{aligned}$$

Since c is arbitrary, we get

$$\begin{aligned} \mathrm{Cov}(\hat{\mu}) &= \mathrm{Cov}(\tilde{\mu}) && \text{if } r = k, \\ &\leqslant \frac{(\tau_{i_1} + \tau_{i_k})^2}{4\tau_{i_1}\tau_{i_k}}\ \mathrm{Cov}(\tilde{\mu}) && \text{if } r < k. \end{aligned} \tag{3.7}$$

The results (3.5)–(3.6) follows from (3.3), (3.4), and (3.7). ∎

Remark. Since it is always true that $\mathscr{M}(B^{1/2}X) \subset \mathscr{M}(\xi_1, ..., \xi_n)$, a special case of (3.5) is the following unconstrained Kantorovich inequality:

$$X'BA^{-1}BX \leqslant \frac{(\tau_1 + \tau_n)^2}{4\tau_1\tau_n}\ X'BX(X'AX)^{-}\ X'BX.$$

The following corollaries are some interesting special cases of Theorem 2. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ be eigenvalues of a positive definite matrix A and $\varphi_1, ..., \varphi_n$ be the corresponding orthonormal eigenvectors.

COROLLARY 3. *Suppose that* $\mathscr{M}(X) \subset \mathscr{M}(\varphi_{i_1}, ..., \varphi_{i_k})$ *for some* $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$. *Then*

$$X'A^{-1}X \leqslant \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} X'X(X'AX)^- X'X. \tag{3.8}$$

If $(X'X)^{-1}$ *exists, then*

$$(X'X)^{-1} X'A^{-1}X(X'X)^{-1} \leqslant \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} (X'AX)^{-1}. \tag{3.9}$$

In particular, if $X'X = I_p$, *then*

$$X'A^{-1}X \leqslant \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} (X'AX)^{-1}. \tag{3.10}$$

COROLLARY 4. *Suppose that* $\mathscr{M}(A^{1/2}X) \subset \mathscr{M}(\varphi_{i_1}, ..., \varphi_{i_k})$, $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$. *Then*

$$X'A^2X \leqslant \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} (X'AP_XAX), \tag{3.11}$$

where $P_X = X(X'X)^- X'$. *If* $X'X = I_p$, *then*

$$X'A^2X \leqslant \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} (X'AX)^2. \tag{3.12}$$

COROLLARY 5. *Let* A *and* A^{-1} *be partitioned as*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \qquad A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix},$$

where A_{11} *is* $m \times m$ *and* A_{22} *is* $(n-m) \times (n-m)$. *Suppose that* $\mathscr{M}(Z_m) \subset \mathscr{M}(\varphi_{i_1}, ..., \varphi_{i_k})$, $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$, *where* $Z_m = (I_m, 0)'$ *is an* $n \times m$ *matrix. Then*

$$A^{11} \leqslant \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} A_{11}^{-1}. \tag{3.13}$$

Proof. Result (3.13) follows from (3.10) by letting $X = Z_m$. ∎

Special cases of (3.8)–(3.13) are the unconstrained Kantorovich inequalities obtained by letting $\lambda_{i_1} = \lambda_1$, $\lambda_{i_k} = \lambda_n$, since $\mathscr{M}(\varphi_1, ..., \varphi_n)$ is the whole space. The unconstrained version of (3.10) and (3.13) are also given by Marshall and Olkin [9] and their extensions to the case of singular A are provided by Baksalary and Puntanen [1].

The inequalities (3.5), (3.8)–(3.13) have many statistical applications. We discuss two examples.

EXAMPLE 5. Consider the following linear model

$$y = X\beta + \varepsilon, \qquad E(\varepsilon) = 0, \qquad \mathrm{Cov}(\varepsilon) = \sigma^2 I_n, \tag{3.14}$$

where X is $n \times p$ and β is $p \times 1$. Suppose that $\beta = (\beta_1', \beta_2')'$, where β_1 is $m \times 1$ and β_2 is $(p-m) \times 1$, and $X = (X_1, X_2)$, where X_1 is $n \times m$ and X_2 is $n \times (p-m)$. If we know $\beta_2 = 0$, then β_1 can be estimated by

$$\tilde{\beta}_1 = (X_1' X_1)^{-1} X_1' y,$$

which is the BLUE since $\mathrm{Cov}(\varepsilon) = \sigma^2 I_n$. However, $\tilde{\beta}_1$ is not robust against the violation of the assumption of $\beta_2 = 0$. The LSE $\hat{\beta}_1$ under model (3.14) is robust but not efficient when $\beta_2 = 0$ is true. Hence it is of interest to study the relative efficiency of $\hat{\beta}_1$ under $\beta_2 = 0$. When $\beta_2 = 0$,

$$\mathrm{Cov}(\tilde{\beta}_1) = \sigma^2 (X_1' X_1)^{-1}$$

and

$$\mathrm{Cov}(\hat{\beta}_1) = \sigma^2 (X'X)^{11},$$

where $(X'X)^{11}$ is the upper left $m \times m$ submatrix of $(X'X)^{-1}$. It follows from Corollary 5 that if $\mathcal{M}(Z_m) \subset \mathcal{M}(\varphi_{i_1}, ..., \varphi_{i_k})$, $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$, then

$$\mathrm{Cov}(\hat{\beta}_1) \leqslant \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} \mathrm{Cov}(\tilde{\beta}_1),$$

where $Z_m = (I_m, 0)'$ and $\varphi_1, ..., \varphi_n$ are orthonormal eigenvectors of $X'X$ corresponding to eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. In particular, since $\mathcal{M}(Z_m) \subset \mathcal{M}(\varphi_1, ..., \varphi_n)$, we have

$$\mathrm{Cov}(\hat{\beta}_1) \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n} \mathrm{Cov}(\tilde{\beta}_1).$$

EXAMPLE 6. We now apply Theorem 2 to obtain an upper bound for the asymptotic covariance matrix of the weighted least squares estimator in the generalized linear model (GLM). The GLM is a natural extension of the linear model (2.2). In the GLM, the relation between the observation y_i and the explanatory variable x_i is not linear but

$$E(y_i) = \mu(\theta_i), \qquad \mathrm{Var}(y_i) = \phi\dot{\mu}(\theta_i), \qquad i = 1, ..., n,$$

where μ is a function on R and $\dot{\mu}$ is its derivative, $g(\mu(\theta_i)) = x_i'\beta$, and g is a known function called the link function, x_i is a $p \times 1$ vector of known values, and β is a $p \times 1$ vector of unknown parameters. No assumption on the joint distribution of $y = (y_1, ..., y_n)'$ is made except that we assume the covariance matrix of y is of the form (2.10), i.e., a block diagonal matrix with small block sizes.

Note that if $\mu(t) \equiv t$ and $g(t) \equiv t$, then the GLM reduces to the linear model (2.2).

The weighted least squares estimator $\hat{\beta}$ of β in the GLM is defined to be a solution of

$$X' \Delta S = 0,$$

where $X = (x_1, ..., x_n)'$, $\Delta = \operatorname{diag}(h(x_1'\beta), ..., h(x_n'\beta))$ with $h(t) = d(g(\mu))^{-1}/dt$, and $S = (y_1 - g^{-1}(x_1'\beta), ..., y_n - g^{-1}(x_n'\beta))'$. Under some regularity conditions, it can be shown that (e.g., Theorem 1 in Liang and Zeger [7]) $\hat{\beta}$ is asymptotically normal with mean β and the asymptotic covariance matrix

$$\Sigma = (X' \Delta \Lambda \Delta X)^{-1} X' \Delta \operatorname{Cov}(y) \Delta X (X' \Delta \Lambda \Delta X)^{-1},$$

where $\Lambda = \operatorname{diag}(\dot{\mu}(\theta_1), ..., \dot{\mu}(\theta_n))$.

Let $B = \Delta \Lambda \Delta$ and $A = \Delta \Lambda [\operatorname{Cov}(y)]^{-1} \Lambda \Delta$. Then it follows from Theorem 2 that if $\mathscr{M}(B^{1/2}X) \subset \mathscr{M}(\varphi_{i_1}, ..., \varphi_{i_k})$, $1 \leqslant i_1 \leqslant \cdots \leqslant i_k \leqslant n$, then

$$\Sigma \leqslant \frac{(\lambda_{i_1} + \lambda_{i_k})^2}{4\lambda_{i_1}\lambda_{i_k}} \{X' \Delta \Lambda [\operatorname{Cov}(y)]^{-1} \Lambda \Delta X\}^{-1},$$

where $\varphi_1, ..., \varphi_n$ are orthonormal eigenvectors of $\operatorname{Cov}(y)$ corresponding to the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. In particular,

$$\Sigma \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} \{X' \Delta \Lambda [\operatorname{Cov}(y)]^{-1} \Lambda \Delta X\}^{-1}.$$

References

[1] Baksalary, J. K., and Puntanen, S. (1991). Generalized matrix versions of the Cauchy–Schwarz and Kantorovich inequalities. *Aequationes Math.* **41** 103–110.

[2] Blom, G. (1976). When is the arithmetic mean BLUE? *Amer. Statist.* **30** 40–42.

[3] Bloomfield, P., and Watson, G. S. (1975). The efficiency of least squares. *Biometrika* **62** 121–128.

[4] Khatri, C. G., and Rao, C. R. (1981). Some extensions of the Kantorovich inequality and statistical applications. *J. Multivariate Anal.* **11** 498–505.

[5] Khatri, C. G., and Rao, C. R. (1982). Some generalizations of Kantorovich inequality. *Sankhyā* **44** 91–102.

[6] Knott, M. (1975). On the minimum efficiency of least squares. *Biometrika* **62** 129–132.

[7] Liang, K. Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22.

[8] Magness, T. A., and McGuire, J. B. (1962). Comparison of least squares and minimum variance estimates of regression parameters. *Ann. Math. Statist* **33** 462–470.

[9] Marshall, A. W., and Olkin, I. (1990). Matrix versions of the Cauchy and Kantorovich inequalities. *Aequationes Math.* **40** 89–93.

[10] Rao, C. R. (1985). The inefficiency of least squares: Extensions of the Kantorovich inequality. *Linear Algebra Appl.* **70** 249–255.

[11] Rao, C. R., and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications.* Wiley, New York.

[12] Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092–1109.

[13] Zyskind, G. (1969). Parameter augmentations and error structures under which certain simple least squares and analysis of variance procedure are also best. *J. Amer. Statist. Assoc.* **64** 1353–1368.